Willingness to Pay for a Quality-Adjusted Life Year: Implications for Societal Health Care Resource Allocation

Joseph T. King, Jr., MD, MSCE, Joel Tsevat, MD, MPH, Judith R. Lave, PhD, Mark S. Roberts, MD, MPP

Background. *Health-state preferences can be combined with willingness-to-pay (WTP) data to calculate WTP per quality-adjusted life year (QALY). The WTP/QALY ratios provide insight into societal valuations of expenditures for medical interventions.* ***Methods***. *The authors measured preferences for current health in 3 patient populations (N = 391) using standard gamble, time trade-off, visual analog scale, and WTP, then they calculated WTP/QALY ratios. The ratios were compared with several proposed cost/QALY cost-effectiveness ratio thresholds, the value-of-life literature, and with WTP/QALY ratios derived from published preference research.* ***Results***. *Mean WTP/QALY ratios ranged from $12,500 to $32,200 (2003 $US). All values were below most published cost-effectiveness ratio thresholds, below the ratio from a prototypic medical treatment covered by Medicare (i.e., renal dialysis), and below ratios from the value-of-life literature. The WTP/QALY ratios were similar to those calculated from published preference data for patients with symptomatic menopause, dentofacial deformities, asthma, or dermatologic disorders.* ***Conclusions***. *WTP/QALY ratios calculated using preference data collected from diverse populations are lower than most proposed thresholds for determining what is "cost-effective." Current proposed cost-effectiveness ratio thresholds may overestimate the willingness of society to pay for medical interventions.* ***Key words:*** *cost-effectiveness analysis; preferences; QALYs; utility; willingness to pay.* ***(Med Decis Making 2005;25:667–677)***

Cost-effectiveness analysis (CEA) combines survival, health-related quality of life, and resource consumption into a single outcomes metric, the incremental cost-effectiveness ratio (ICER). ICERs can be used to evaluate health care interventions and to determine whether they are reasonable expenditures of limited resources.[1] The quality-of-life values used in CEA are generally measured on a ratio scale anchored on death and perfect health using the standard gamble[2] (SG) or time trade-off[3] (TTO). These techniques differ with respect to several characteristics: the SG incorporates attitudes toward risk of death, and the TTO involves attitudes toward certain loss of survival time. The visual analog scale (VAS) is also used to elicit quality of life[4]; however, VAS scores are not grounded in expected utility theory and consequently should not be substituted for utilities derived from the SG or TTO.

Several methods have been suggested to evaluate the ICER of an intervention: 1) league tables—a ranking of ICERs for various interventions; 2) ICER thresholds—specific incremental ICER values that demarcate the limits of cost effectiveness; and 3) societal willingness to pay (WTP)—the hypothetical limits to resources that society is willing to allocate to medical interventions. These evaluation techniques each have their shortcomings, and they can produce dissimilar results. League tables are dependent on several factors, including the

Received 18 November 2003 from the VA Connecticut Healthcare System, West Haven, CT (JTK); the Department of Neurosurgery, Yale University, New Haven, CT (JTK); the Department of Internal Medicine and the Center for Clinical Effectiveness at the Institute for Health Policy and Health Services Research, University of Cincinnati Medical Center, Cincinnati, OH (JT); the Veterans Affairs Medical Center, Cincinnati, OH (JT); the Department of Health Policy and Management, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA (JRL); the Center for Research on Health Care, University of Pittsburgh, Pittsburgh, PA (JRL, MSR); and the Department of Medicine and the Division of General Internal Medicine, University of Pittsburgh, Pittsburgh, PA (MSR). Dr. King is supported by a Veterans Administration VISN 4 CPPF grant and by a Mentored Patient-Oriented Research Career Award from the National Institute of Neurological Disorders and Stroke (1K23 NS02169-06). Revision accepted for publication 7 September 2005.

DOI: 10.1177/0272989X05282640

comparator studies.[5] Benchmark ICER thresholds have been proposed by Laupacis and colleagues,[6] Kaplan and Bush,[7] and the Development and Evaluation Committee of the South and West Regional Office of the National Health Service (NHS) in the United Kingdom[8]; however, the proposed thresholds differ significantly, particularly when adjusted for inflation and currency exchange rates. Societal WTP can be assessed by several different techniques: the value-of-life literature, health care resource allocation policy (e.g., Medicare funding of renal dialysis), and WTP per quality-adjusted life year (QALY) ratios based on empiric data.

Rather than using arbitrary decision rules such as league tables or ICER thresholds, it may be more reasonable to allocate health care resources based on societal WTP for health care benefits. Societal WTP can be obtained by examining the value-of-life literature, societal health care resource allocation policies such as the public funding of renal dialysis by Medicare in the US, and studies of health-state values that use both utility and contingent valuation methodologies. The value-of-life literature suggests 3 methods to value a human life: revealed preference, human capital, and contingent valuation.[9] In revealed preference, one can infer the value of life by analyzing the real-world actions of individuals who are willing to face risk for payment (e.g., hazard pay for dangerous occupations) or who are willing to pay to reduce risk (e.g., smoke detectors to reduce the risk from fire). The human capital approach posits that the value of saving a life is equivalent to the potential additional cumulative lifetime earnings resulting from longer survival. Contingent valuation, also known as WTP, is based on a hypothetical market that uses the price an individual is willing to pay to obtain a beneficial intervention.[10] Johannesson and Meltzer have recommended that investigations of societal WTP for health care benefits should be a "research priority."[5]

In this study, we report WTP/QALY ratios obtained from a population of general medical patients and from 2 neurosurgical patient populations—one afflicted with cervical spondylotic myelopathy (CSM), a degenerative spine condition, and the other harboring cerebral aneurysms. We then compare these values to proposed ICER thresholds, several societal WTP standards, and WTP/QALY ratios derived from studies of utility and WTP in patients with specific medical conditions. Finally, we discuss the implications of our findings for the allocation of health care resources.

METHODS

Study Populations

We studied 3 groups of patients: 1) veterans diagnosed with CSM recruited from October 2000 to September 2001 from the Veterans Affairs Pittsburgh Healthcare System (VAPHS) neurosurgery clinic, 2) a convenience sample of veterans recruited from January 2001 to December 2001 from VAPHS general medical clinics, and 3) patients with cerebral aneurysms recruited from June 2001 to February 2004 from neurosurgery clinics at the University of Pittsburgh Medical Center (UPMC). The current study is a supplementary analysis of data collected in separate studies of preferences in patients with neurosurgical conditions and general medical clinic patients.[11–13] Informed consent was obtained from all subjects prior to data acquisition, and the protocols were approved by the institutional review boards of the VAPHS (#00010) and the University of Pittsburgh (#000936 and #0002117). Subjects underwent a structured interview administered by trained research assistants collecting data on demographics, income, preferences, and health values, and they received $25 at the completion of the data collection process.

Preference Testing

We used the VAS, SG, TTO, and WTP to measure subjects' preferences for their current health. The anchor points for testing were death and perfect health, the latter defined as "the best possible health that you can imagine." Preference testing was administered by a research assistant using a script and interactive graphical software running on a portable computer. An iMPACT3 software program[14] was used for SG and TTO testing in all study populations and for VAS testing in patients with CSM and general medical clinic patients. A pencil-and-paper instrument was used for VAS testing in patients with aneurysms. WTP was assessed with a custom Visual Basic program in all study populations. The order of preference testing was VAS, SG, TTO, and WTP, and all subjects were tested with each instrument.

Visual Analog Scale. Subjects were presented with a vertical "thermometer" image on the computer screen

In the current article, inflation adjustments used a ratio of price valuations between 2003 and the year that the data were collected, and the currency conversion rates varied with the native currency of the primary data source being used for the WTP/QALY calculations.

Address correspondence to Joseph T. King, Jr., MD, MSCE, Section of Neurosurgery, VA Connecticut Healthcare System/112, 950 Campbell Avenue, West Haven, CT 06516; e-mail: Joseph.KingJr@med.va.gov.

(patients with CSM and general medical clinic patients) or with a 10-cm line printed on a piece of paper (patients with aneurysms). The thermometer or line was anchored at the lower end by death and at the upper end by perfect health. Subjects indicated the value of their current health by adjusting the level of the thermometer or by inscribing a mark on the line. The value of current health was calculated as the exact number on the thermometer, or on the 10-cm line as the distance from death to the mark for current health divided by the total length of the line.

Standard Gamble. Subjects were offered a choice between continuing to live in their current state of health or accepting a hypothetical treatment for all of their medical problems and symptoms. The treatment had 2 possible outcomes: death or perfect health. The probabilities of death and perfect health were varied systematically using a ping-pong technique until the subject was indifferent between current health and the treatment results. The utility was equivalent to the probability of perfect health at the indifference point.

Time Trade-off. Subjects were offered a choice between continuing to live in their current state of health or trading years of life in exchange for an immediate permanent cure of all medical problems and symptoms. The number of years traded off to obtain perfect health was varied systematically using a ping-pong technique until the subject was indifferent between current health and the trade-off. All subjects were presented with a 20-year life expectancy. The utility was equivalent to the ratio between time in perfect health and time in current health at the indifference point.

Willingness to Pay. We used an iterative closed-ended bidding method to determine WTP for a hypothetical treatment resulting in perfect health by curing of all of the subjects' health problems and symptoms. Subjects were asked to imagine that they could purchase a cure for all of their health problems and symptoms with a single payment. The initial price was $1, the 2nd price offered was equivalent to 1 month of their household income, and the maximum price permitted was 10 times the subject's own annual household income. A computer program calculated each successive price offer using a bisecting algorithm incorporating annual household income and the subject's last response. When patients refused to provide their annual household income, a proxy value ($25,000 for patients with CSM or general medical clinic patients, $35,000 for patients with aneurysms) was used during WTP assessments. The WTP value was the maximum price that the subject would pay to obtain perfect health.

Health Status

CSM and general medical subjects completed the Medical Outcomes Study-Short Form 36 item survey (SF-36).[15] Aneurysm patients completed the Medical Outcomes Study-Short Form 12 item survey (SF-12).[16] Survey responses were used to calculate physical component summary (PCS) and mental component summary (MCS) scores using standard algorithms.[17] The PCS and MCS scores range from 0 to 100, with a mean of 50 and a standard deviation of 10; higher values represent better health. PCS and MCS scores from the SF-36 and SF-12 are highly correlated and can be compared.[16]

Data Analysis

Categorical variables were tabulated, and medians, quartiles, means, and standard deviations were calculated for continuous variables. We used VAS, SG, and TTO results for current health to estimate cumulative lifetime QALYs being "purchased" by multiplying the difference between current health and perfect health by life expectancy, as garnered from US Life Tables.[18] The US Consumer Price Index (CPI) was used to inflate monetary values to 2003 US dollars.[19] We used a 3% discount rate to calculate the present value of future health benefits[1]; the 1-time WTP payment format obviated the need to discount future costs. We calculated the WTP/QALY ratio as follows:

$$WTP\ /\ QALY = \frac{\textit{Willingness to pay amount}}{\sum_{t=1}^{\textit{life}\ \text{expectancy}} (1 - \textit{value of current health}) * (1 + \textit{discount rate})^{-(t-1)}}.$$

Stepwise multivariate linear regression was used to model the association between WTP/QALY calculated with the VAS, SG, or TTO and patient characteristics. For regression modeling, we created a combined data set ($N = 391$) containing data from patients with CSM, general medical clinic patients, and patients with cerebral aneurysms. Simple linear regression and a threshold $P < 0.2$ were used to select candidate-predictor variables for inclusion in the stepwise multivariate models. Predictor variables assessed included the following: clinic site, age, sex, race, self-reported disability status, employment status, retirement status, income, PCS, and MCS. The distribution of income was normalized using a natural log transformation. In the final multivariate models, $P < 0.05$ was considered sta-

Table 1 Characteristics of the Study Populations

	Patients with CSM (n = 84)	General Medical Clinic Patients (n = 117)	Patients with Cerebral Aneurysms (n = 190)	All Patients (N = 391)
Age (y)				
$\bar{x}$ [s]	57.1 [11.2]	59.8 [13.6]	54.4 [12.6]	56.6 [12.8]
Range	29–84	24–84	25–90	24–90
Male gender (n [%])	77 [92]	105 [90]	51 [27]	233 [60]
Race (n [%])				
Non-Hispanic White	73 [87]	87 [74]	174 [92]	334 [85]
African American	5 [6]	24 [21]	14 [7]	43 [11]
Hispanic	1 [1]	0 [0]	1 [0.5]	2 [0.5]
Native American	4 [5]	4 [3]	0 [0]	8 [2]
Other	1 [1]	2 [2]	0 [0]	3 [0.8]
Declined to answer	0 [0]	0 [0]	1 [0.5]	1 [0.3]
Income (2003 $US)				
Median [1st, 3rd quartiles]	16,000 [15,600, 26,700]	15,600 [15,600, 26,000]	35,000 [15,600, 46,800]	26,000 [15,600, 36,400]
Declined to answer (n [%])	2 [2]	6 [5]	15 [8]	23 [4]
Disabled				
n [%]	58 [69]	63 [54]	57 [30]	178 [46]
Declined to answer	4 [5]	0 [0]	3 [2]	7 [2]
Currently employed				
n [%]	14 [17]	32 [27]	89 [47]	135 [35]
Declined to answer	4 [5]	0 [0]	2 [1]	6 [2]
Retired				
n [%]	37 [44]	70 [60]	52 [27]	159 [41]
Declined to answer	4 [5]	0 [0]	5 [3]	9 [2]
SF-12 ($\bar{x}$ [s])				
PCS score	28.1 [8.4]	37.1 [11.7]	40.0 [9.4]	36.5 [11.0]
MCS score	40.8 [12.8]	45.3 [14.1]	39.3 [7.8]	41.4 [11.5]

Note: CSM = cervical spondylotic myelopathy; s = standard deviation; SF-12 = Medical Outcomes Study-Short Form 12; PCS = physical component summary; MCS = mental component summary.

tistically significant, and $0.05 \le P < 0.10$ was considered a trend. Since our study populations had relatively low incomes, we repeated our analyses on a random sample of the combined data set stratified to match the income distribution of the US population in 2002.[20]

RESULTS

Eighty-four patients with CSM, 117 general medical clinic patients, and 190 patients with cerebral aneurysms completed all assessments (N = 391) (Table 1). Consistent with the demographics of veterans and the epidemiology of cerebral aneurysms, the patients with CSM and the general medical patients were predominantly men (≥90%), and the majority of patients with cerebral aneurysms were women (73%). Nonwhites constituted 26% or less of each of the 3 study populations. The median annual household income was approximately $16,000 for both veteran population samples and $35,000 for the UPMC aneurysm population sample.

Preference testing results for the value of current health varied across the 3 populations (Table 2). Patients with CSM had the lowest mean values for current health: VAS, 0.51; SG, 0.72; and TTO, 0.71. General medical patients had intermediate values for current health: VAS, 0.64; SG, 0.76; and TTO, 0.77. Patients with cerebral aneurysms had the highest values for current health: VAS, 0.67; SG, 0.77; and TTO, 0.79. Mean WTP values were higher in patients with aneurysms ($119,900) compared with patients with CSM ($104,900) and general medical patients ($83,400). Combining the utility and WTP responses into WTP/

Table 2 Preferences for Current Health State and Calculated WTP/QALY

Population	n	Preferences, $\bar{x}$ (s)				WTP/QALY,[a] $\bar{x}$		
		VAS	Standard Gamble	Time Trade-off	Willingness to Pay[a]	VAS	Standard Gamble	Time Trade-off
Patients with CSM	84	0.51 (0.19)	0.72 (0.24)	0.71 (0.28)	\$104,900 (\$141,100)	\$12,500	\$22,400	\$22,700
General medical clinic patients	117	0.64 (0.22)	0.76 (0.22)	0.77 (0.27)	\$83,400 (\$150,700)	\$14,700	\$21,000	\$22,900
Patients with cerebral aneurysms	190	0.67 (0.20)	0.77 (0.24)	0.79 (0.25)	\$119,900 (\$186,600)	\$32,200	\$29,500	\$32,200
All patients	391	0.63 (0.21)	0.76 (0.23)	0.76 (0.26)	\$105,800 (\$167,700)	\$16,600	\$25,400	\$27,100

Note: WTP/QALY = willingness to pay per quality-adjusted life year; VAS = visual analog scale; CSM = cervical spondylotic myelopathy; s = standard deviation.
a. Adjusted to 2003 US dollars; see Methods section for calculation details.

Table 3 Regression Models of WTP/QALY

Dependent Variable	Regression Coefficients							
	Income (2003 \$US)[a]	Age	Retired	PCS Score	MCS Score	Constant	F	Model R^2
WTP/QALY using VAS	53,656[e]		71,694[c]	2524[d]		–611,143[e]	0.002	0.04
WTP/QALY using standard gamble	92,542[b]					–858,980[b]	<0.001	0.05
WTP/QALY using time trade-off	31,817[b]	820[c]			–688[d]	–302,910[b]	<0.001	0.15

Note: WTP/QALY = willingness to pay per quality-adjusted life year; PCS = physical component summary; MCS = physical component summary; VAS = visual analog scale; s = standard deviation.
a. Natural log transformation.
b. $P < 0.001$.
c. $P < 0.05$.
d. $P < 0.1$.
e. $P < 0.01$.

QALY ratios (Table 2), all 3 study populations had ratios between \$12,500 and \$32,200.

Regression models predicting WTP/QALY calculated with the VAS, SG, and TTO showed that patient income was consistently associated with WTP/QALY ratios—wealthier patients had higher WTP/QALY ratios (Table 3). Although the model R^2 values were low, the magnitude of the income effect was substantial. For example, in the SG model, an income change from \$25,000 to \$50,000 would increase the predicted WTP/QALY ratio by \$64,100. Greater age (TTO), lower MCS scores (TTO), higher PCS scores (VAS), and retirement (VAS) were associated with higher WTP/QALY ratios (Table 3).

The stratified random sample of 62 subjects had a median income of \$46,800, similar to the US 2002 median household income of \$43,200. The mean subgroup WTP of \$166,800 was approximately 60% greater than the \$105,800 WTP for the combined data set. The corresponding WTP/QALY ratios for the upper-income subgroup were \$26,000 for VAS, \$43,000 for SG, and \$43,400 for TTO.

DISCUSSION

The decision to devote resources to a health care intervention implies that the value of the anticipated health benefit from the intervention should be equal to or greater than the cost of the intervention. Cost-benefit analysis makes this explicit determination by assigning monetary values to all of the intervention-associated costs and to all of the resulting benefits. If the monetary value of the benefits exceeds the costs, then the intervention has a net benefit and should be implemented. Because the monetary valuation of many health care benefits is quite difficult and controversial, cost-benefit analyses are relatively rare in the biomedical literature. Cost-effectiveness analysis does not require monetary valuation of health benefits, only the calculation of a ratio of net costs to net health effects, an easier task. The output of a cost-effectiveness analysis is the ICER, the incremental cost required to produce an incremental increase in health, commonly expressed as cost/QALY. Several approaches such as league tables, absolute thresholds, and determinations of societal WTP have been developed to determine whether a particular ICER justifies the resource expenditure.

Threshold Values

Threshold values are appealing in their simplicity, but they vary widely with the source (Table 4). Kaplan and Bush proposed $50,000/QALY (1982 $US) as a threshold value for determining whether an intervention was cost effective.[7] Using the US CPI[19] to adjust for inflation, this threshold is equivalent to $95,000/QALY in 2003 US dollars. (Despite the obvious impact of inflation on the 20-year-old figure, many investigators continue to quote the original, unadjusted $50,000/QALY threshold.[21]) Laupacis and colleagues proposed a 3-tiered standard ostensibly based on empiric data, citing 1) "strong" evidence for adopting an intervention if its ICER is <$20,000/QALY, 2) "moderate" evidence for adopting an intervention if its ICER range is $20,000–$100,000/QALY, and 3) "weak" evidence for adopting an intervention if its ICER is >$100,000/QALY (1990 $CAN).[6] The authors noted that "the nominal figures should be adjusted periodically to maintain constant value in real terms (adjusted for increases in the price level)."[6(p476)] Adjusting for both inflation and currency exchange rates, these values are the equivalent of <$23,400/QALY, $23,400–$116,800/QALY, and >$116,800/QALY, respectively, in 2003 $US. The Development and Evaluation Committee of the South and West Regional Office of the NHS in the United Kingdom incorporates explicit thresholds of cost-effectiveness ratios into its evaluations of health technology. Adjusted for inflation and international exchange rates, cost-effectiveness ratios are categorized as follows: A, <$5200/QALY; B, $5200–$34,700/QALY; C, >$34,700/QALY; and D, negative life years (2003 $US). The Committee combines the cost-effectiveness data (categories A–D) and an evaluation of the quality of the evidence supporting the cost-effectiveness data to categorize medical interventions as *strongly recommended*, *recommended*, *borderline*, *not recommended*, or *not proven*. The categories are then used to allocate NHS health care resources.

Value-of-Life Literature

Hirth and colleagues recently published an extensive review and analysis of the value-of-life literature.[9] The authors performed a secondary data analysis of 42 published articles, calculating the WTP/QALY using monetary values contained in the report, study population life expectancy (estimated from age and life tables), study population quality-of-life weights (estimated from age and Beaver Dam population normative values[22]), and discounted future costs and benefits. They found a wide range of values for WTP/QALY across the value-of-life literature (Table 4). Median values for WTP/QALY varied significantly by the valuation methodology of the studies: human capital studies, $28,300; contingent valuation studies, $184,200; studies revealing preference derived from job risk, $381,500; and studies revealing preference derived from safety, $106,700 (all values adjusted to 2003 $US).

"Dialysis Standard"

Some have argued that public funding in the United States for renal dialysis through the Medicare entitlement program establishes a de facto societal threshold for WTP for a health care intervention.[23,24] Others dispute this interpretation, noting that renal dialysis is a unique case of universal public funding for the treatment of a particular disease and that it is not appropriate to extrapolate a societal standard from this lone legislative fiat. A recent estimate placed the ICER of renal dialysis at $74,000–$95,000/QALY (1997 $US), the equivalent of $84,500–$108,500/QALY in 2003 US dollars[9] (Table 4).

ICER Limitations

The use of ICERs to evaluate health interventions has been criticized for utilizing a limited perspective that does not reflect the economic realities of health policy decisions. Gafni and Birsh argued that adopting more effective yet more expensive health technologies, even if the new technology has a favorable ICER, leads to increased net health expenditures.[25] Diversion of resources from other health interventions or from nonhealth expenditures may be necessary to pay for the new intervention, and the "opportunity cost" of these expenditures should be included in any consideration of ICERs. For example, a policy maker might

Table 4 Standards for Evaluating WTP/QALY Ratios

Source	Year Published	Country	Description	Benchmark (WTP/QALY)[a]
Kaplan and Bush[7]	1982	US	Proposed standard	$95,000
Laupacis et al.[6]	1992	Canada	Proposed standard—strong evidence for adoption	<$23,400
			Proposed standard—moderate evidence for adoption	$23,400–$116,800
			Proposed standard—weak evidence for adoption	>$116,800
Hirth et al.[9]	2000	US	"Dialysis standard"—lower limit	$84,500
			"Dialysis standard"—upper limit	$108,500
Hirth et al.[9]	2000	Several	Value-of-life literature, human capital method (median value, 6 studies)	$28,300
			Value-of-life literature, revealed-preference method using risky-occupation data (median value, 19 studies)	$381,500
			Value-of-life literature, revealed-preference method using willingness-to-pay-for-safety data (median value, 8 studies)	$106,700
			Value-of-life literature, contingent-valuation method (median value, 8 studies)	$184,200
Woolf and Henshall[8]	2000	UK	Actual standard—category A[b]	<$5200
			Actual standard—category B[b]	$5200–$34,700
			Actual standard—category C[b]	>$34,700
			Actual standard—category D[b]	Negative life years

a. Willingness to pay per quality-adjusted life year (WTP/QALY) adjusted to 2003 US dollars.
b. The Development and Evaluation Committee of the South and West Regional Office of the National Health Service in the United Kingdom uses cost-effectiveness data (categories A–D) and an evaluation of the quality of the evidence supporting the cost-effectiveness data to categorize medical interventions as *strongly recommended*, *recommended*, *borderline*, *not recommended*, or *not proven*. The categories are then used to allocate health care resources.

implement a new "cost-effective" intervention by identifying and eliminating a less cost-effective and similarly priced intervention from the budget, thus increasing health benefits while keeping expenditures constant.

WTP/QALY Values in the Biomedical Literature

Johannesson and Meltzer have noted that it is possible to calculate WTP/QALY by using utilities and WTP measured simultaneously.[5] A few authors have used this approach and have reported the WTP/QALY ratio in their original publications. Zethraeus examined quality of life using the TTO, VAS, and monthly WTP payments for 3 years of hormone replacement therapy in symptomatic menopausal women from Sweden, calculating the WTP/QALY at 120,000 to 160,000 SK (1995–1996 SK), the equivalent of $21,200 to $28,300 (2003 $US).[26] Several other authors have reported both utility and WTP for current health, but they did not calculate the WTP/QALY ratio. One can calculate WTP/QALY from their data by 1) determining the time horizon of the health state under study (either as explicitly stated in the study, or estimated from the study population age and life tables); 2) extrapolating cumulative payments from the WTP data and the time horizon; 3) adjusting monetary values for inflation[19] and for foreign currency exchange rates[27] (if applicable); 4) extrapolating the cumulative quality-adjusted survival difference between current health and perfect health over the time horizon; 5) discounting future payments and quality-adjusted survival at 3%[1]; and 6) dividing total discounted payments by total discounted QALYs (Table 5). For example, Cunningham and Hunt used the SG and a single WTP payment to measure preferences for current health in a population of British patients with dentofacial deformities, yielding a WTP/QALY of $1800 (2003 $US).[28] Lundberg and others used the SG, TTO, and perpetual monthly WTP payments to measure preferences for current health in Swedish patients with psoriasis or eczema.[29] They found wide variation in the WTP/QALY ratios depending on the utility assessment method: patients with psoriasis had WTP/QALY ratios of $94,700 using the SG and $23,700 using the TTO; patients with atopic eczema had WTP/QALY ratios of $94,700 using the SG and $27,100 using the TTO (all 2003 $US). Blumenschein and Johannesson used the SG, TTO, and monthly WTP payments to mea-

Table 5 WTP/QALY Ratios Derived from the Published Medical Literature

					WTP/QALY[a]		
Source	Year Published	Data Collection	Country	Condition	VAS	Standard Gamble	Time Trade-off
Blumenschein and Johannesson[30]	1997	1995–1996	US	Asthma	$9400	$33,400	$23,600
Zethraeus[26]	1998	1995–1996	Sweden	Symptomatic menopause	$20,900	—	$27,600
Lundberg et al.[29]	1999	1996–1997	Sweden	Psoriasis	$9200	$94,700	$23,700
				Atopic eczema	$7000	$94,700	$27,100
Cunningham and Hunt[28]	2000	1998	UK	Dentofacial deformity	—	$1800	—

Note: WTP/QALY = willingness to pay per quality-adjusted life year; VAS = visual analog scale.
a. Adjusted to 2003 US dollars.

sure preferences for current health in patients from the United States with asthma.[30] The WTP/QALY ratios varied with the utility assessment method: SG, $33,400 and TTO, $23,600 (all 2003 $US).

Current Study

The WTP/QALY ratios in our neurosurgical and general medical study populations ranged from $12,500 to $32,200 and were similar to most values in the medical literature from patients with asthma, dermatologic disorders, dentofacial deformities, or menopausal symptoms. Strikingly, the WTP/QALY ratios from our subjects and ratios derived from the published medical literature are all lower than the inflation-adjusted ICER standards of Kaplan and Bush, the upper limits of Laupacis and colleagues' 3-tiered standard, the dialysis standard, and most ratios derived from the value-of-life literature. There were only 2 standards that were not considerably above our values: WTP/QALY ratios calculated from value-of-life studies using the human capital method, and the NHS standards. The human capital method systematically underestimates the benefit of a health intervention because it only considers the impact of the intervention on potential future earnings (i.e., years spent in retirement are given no value). Most economists would not accept this as a reasonable standard—it is not a WTP assessment. A more notable exception is the NHS standard, which differs from the proposed or inferred standards because it is an explicit health policy actually being used to allocate health care resources.

There appears to be a discrepancy between the higher WTP/QALY values based on ICER thresholds or the value-of-life literature, and the lower WTP/QALY values derived from studies of specific medical conditions. With the exception of the SG-based values in the study of Lundberg and coworkers, all of the ratios derived from published studies of specific medical conditions are well below the $50,000 threshold of Kaplan and Bush, near the lower boundary of the middle tier of Laupacis and others, below the dialysis standard, or below the value-of-life literature calculations of Hirth and colleagues. It is worth noting that the value-of-life literature is based on extrapolations of actual human behavior under conditions in which individuals make real choices that affect their survival, such as taking a job with a relatively high risk of death in return for higher income. On one hand, such data may be more accurate than values derived from hypothetical decisions, such as the SG or TTO. On the other hand, decisions regarding higher pay versus a riskier job involve the calculus of low-probability events. Individuals have difficulty accurately incorporating low probabilities into decision making,[31] and slight differences or errors in calculation can produce wide swings in the resultant ratios.

Willingness to Pay

The estimation of WTP is subject to considerable variability; it can be assessed using several different methods that may yield different results, including open-ended questions, bidding games, payment cards, single "take-it-or-leave-it" questions, and "take-it-or-leave-it" questions with follow-up.[10] WTP can also be influenced by income and national standards of living. In our 3 study populations, a natural log transformation of income was the best predictor of WTP/QALY across the VAS, SG, and TTO (although income only explained a small amount of the variation in WTP/QALY ratios). Our study populations had median incomes below that of the US population, and the regression models predicted that this would decrease WTP/QALY ratios. To further explore the income effect, we examined a stratified randomized subgroup of our subjects selected to approximate the distribution of US household incomes. While the WTP/QALY ratios were greater in this subgroup (ranging from \$26,000–\$43,400 across the 3 study populations), they are still below most benchmarks from Table 4. Even when adjusting for income, it appears that individuals are not willing to pay as much for health care as most standards imply.

WTP values are quite high in our study populations, particularly when considered as a percentage of household income. The mean WTP amounts ranged from \$83,400 to \$119,900, corresponding to 3.0 to 4.6 times mean annual income. Subjects might afford this expense by amortizing the single payment via loans or by liquidating assets. Converting the WTP responses to an equivalent monthly payment provides additional insights. For example, the responses of the patients with cerebral aneurysms (mean values: current health measured with the SG, 0.77; single-payment WTP for perfect health, \$119,900; WTP/QALY, \$29,500; and a 3% discount rate) are the equivalent of WTP \$600 per month in perpetuity, 18% of the mean annual income of \$40,200. In addition, many patients do not pay for health care directly out of pocket but rather indirectly via insurance programs financed through taxes and/or employers, and thus their WTP values may be distorted by a lack of purchasing experience.

Greater income was consistently associated with higher WTP/QALY ratios across all health value measurement techniques, but there were additional factors that sporadically influenced responses. Difference in the characteristics of the measurement tools may explain some of the associations. For example, older subjects had higher WTP/QALY ratios when measured with the TTO. Perhaps the prospect of death looms larger for older individuals, and they are less willing to trade off their remaining life span, resulting in increased TTO values and correspondingly higher WTP/QALY ratios. Other relationships are rather inexplicable: retired patients had significantly higher WTP/QALY ratios calculated from VAS responses, there were trends toward an association between lower MCS scores and ratios calculated with TTO values, and there was a trend toward higher PCS scores and WTP/QALY ratios derived from the VAS.

Calculation of WTP/QALY Ratios

The WTP/QALY ratio can be calculated using health valuation data measured using a variety of methods, each of which can affect the ratio. Calculation of quality-adjusted survival requires health-state valuations on a 0–1 ratio scale, which can be obtained using several methods known to yield different results. Health values can be influenced by risk attitudes (SG) and time preferences (TTO). Our results (Table 2) corroborate published studies (Table 5) that demonstrate the variability of health-state valuations and WTP/QALY ratios produced by various health-state valuation techniques. To cite an extreme example, in the study by Lundberg and colleagues, SG-derived ratios were 4 times larger than TTO-derived ratios and 8 times larger than VAS-derived ratios. While their finding may be explained by the risk-averse attitudes of subjects unwilling to risk death in the SG to cure a skin condition, it highlights the challenges of incorporating health valuations into ratios. Despite the variation in health-state values produced by the VAS, SG, and TTO, there is still a striking similarity of most WTP/QALY ratios across studies, including our present work.

The SG and TTO are generally accepted as providing QALY weights suitable for the calculation of quality-adjusted survival. More controversial are the VAS and multiattribute utility theory techniques such as the Quality of Well-Being Self-Assessed,[32] Health Utilities Index,[33] or EuroQol-5D.[34] While some argue that the VAS produces acceptable values, many researchers are concerned that VAS values violate axioms of utility theory and are thus inappropriate for calculating QALYs. Multiattribute utility theory methods are appealing because they use written surveys to collect data and then use a mathematical function to generate QALY weights based on a societal perspective. While a societal perspective is desirable, the extrapolation of societal values from individual responses using a mathematical function is not a true utility and thus may not be appropriate for the calculation of quality-adjusted survival.

There is no standard method for extrapolating health-state valuations into cumulative lifetime quality-adjusted survival. Our method of calculating WTP/QALY was constrained by a limited data set (preference values collected at 1 point in time, subject age, year, and country of data collection), and the resulting mathematical model assumed constant health values and incorporated annual average international currency exchange rates, annual US CPI changes, life expectancy based on national actuarial data, and a constant 3% discount rate. Different assumptions would have yielded different results and possibly affected our conclusions. The value-of-life literature models required similar types of assumptions to convert raw data into WTP/QALY ratios, and they are just as vulnerable to criticism.

A Societal Standard?

Attempts to derive a societal standard for WTP/QALY from data describing human behavior or preferences yield disparate results, depending on the source of the data. The WTP/QALY ratios derived from specific patient populations may not accurately reflect the attitudes of society. There are known systematic differences between the health-state valuations of patients and of others, and these differences may translate into distorted WTP/QALY ratios. Furthermore, our regression analyses found that income was consistently related to WTP/QALY ratios and that other patient characteristics (age, employment status, health status) were sporadically related to WTP/QALY ratios. Ultimately, societal values would be best assessed by measuring WTP/QALY in a large, random or stratified sample designed to represent society. While normative population data are available for TTO values (i.e., Beaver Dam health outcomes study[22]), there are no analogous values for SG, WTP or for WTP/QALY ratios. Valuations of health states vary directly with the familiarity of the rater with the health state,[35–39] so we would expect individuals rating their own health to provide higher valuations than would societal representatives rating the same health states. However, our analyses are based on ratios of WTP and QALYs calculated from VAS, SG, or TTO health valuations. If the shifts in health values associated with familiarity are similar for both the numerator (i.e., WTP) and denominator (calculated QALYs), then the quotient may be similar to values calculated from societal representatives.

It is premature to conclude that our findings indicate actual societal preferences; however, the similarity of our results to those obtained from Swedish and British patients with nonneurosurgical diseases provides support for the generalizability of our findings. The participants in our study were diverse, encompassing adult men and women aged 24 to 90 years with annual incomes ranging from $5000 to more than $100,000 and afflicted with 1 of 2 neurological conditions or a broad spectrum of conditions treated in a primary care medical clinic. While our study population had good age and gender diversity, nonwhites were underrepresented and incomes were low compared with the general US population. Nevertheless, the WTP/QALY ratios were quite similar across the 3 populations that we studied, supporting an underlying consistent valuation for how much patients are willing to pay to improve health. Adjusting for the income effect left us with similar conclusions.

The WTP/QALY ratios derived from value-of-life studies or preference studies straddle both sides of existing standards of "acceptable" ICERs. The value-of-life literature implies that society is willing to pay much more to obtain health benefits than is suggested by other standards. Conversely, our work and the literature on preferences suggest that current absolute standards may overestimate societal WTP for health care interventions. If the absolute standards overestimate societal WTP for health care interventions, then interventions should be reevaluated in light of the emerging lower societal WTP/QALY standard derived from health-state valuations. The result could be the reclassification of interventions formerly thought to be "cost-effective" as not economically worthwhile. Perhaps we should not be surprised that 3 different methods of evaluating ICERs—absolute thresholds, the value-of-life literature, and WTP/QALY ratios calculated from empiric preference data—yield divergent conclusions. Presently, it is impossible to say which (if any) of the 3 techniques is superior or even more useful. The discrepancies between the biomedical literature (including our current work) and the value-of-life literature merit further investigation.

REFERENCES

1. Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in Health and Medicine. New York: Oxford University Press; 1996.
2. von Neumann J, Morgenstern O. Theory of Games and Economic Behavior. New York: Wiley; 1953.
3. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. Health Serv Res. 1972;7:118–33.
4. Streiner DL, Norman GR. Health Measurement Scales. A Practical Guide to Their Development and Use. New York: Oxford University Press; 1989.
5. Johannesson M, Meltzer D. Some reflections on cost-effectiveness analysis. Health Econ. 1998;7(1):1–7.
6. Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. Can Med Assoc J. 1992;146(2):473–81.
7. Kaplan RM, Bush JW. Health-related quality of life measurement for evaluation research and policy analysis. Health Psychol. 1982;1(1):61–80.
8. Woolf SH, Henshall C. Health technology assessment in the United Kingdom. Int J Technol Assess Health Care. 2000;16:591–625.

9. Hirth RA, Chernew ME, Miller E, Fendrick AM, Weissert WG. Willingness to pay for a quality-adjusted life year: in search of a standard. Med Decis Making. 2000;20(3):332–42.
10. Diener A, O'Brien B, Gafni A. Health care contingent valuation studies: a review and classification of the literature. Health Econ. 1998;7(4):313–26.
11. King JT Jr, Tsevat J, Moossy JJ, Roberts MS. Preference-based quality of life measurement in patients with cervical spondylotic myelopathy. Spine. 2004;29:1271–80.
12. King JT Jr, Styn MA, Tsevat J, Roberts MS. "Perfect health" versus "disease free": the impact of anchor point choice on the measurement of preferences and the calculation of disease-specific disutilities. Med Decis Making. 2003;23(3):212–25.
13. King JT Jr, Tsevat J, Roberts MS. Preference-based quality of life in patients with cerebral aneurysms. Stroke. 2005;36(2):303–9.
14. Lenert LA, Sturley A, Watson ME. iMPACT3: internet-based development and administration of utility elicitation protocols. Med Decis Making. 2002;22(6):464–74.
15. Ware JE, Sherbourne CD. A 36-item short-form health survey (SF-36): conceptual framework and item selection. Med Care. 1992;30:473–83.
16. Ware J Jr, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34:220–33.
17. Ware JE, Kosinski M, Keller SD. SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston (MA): Health Assessment Lab; 1994.
18. Arias E. United States Life Tables, 2000. National Vital Statistics Reports. 2002;51(3):7–8.
19. US Department of Labor, Bureau of Labor Statistics. Consumer Price Index—All Urban Consumers. Available at: http://data.bls.gov/cgi-bin/surveymost. Accessed 5 May 2004.
20. DeNarvas-Walt C, Cleveland RW, Webster BH Jr. Income in the United States: 2002. Current Population Reports—Consumer Income. Washington (DC): US Census Bureau; 2003.
21. Ubel PA. What is the price of life and why doesn't it increase at the rate of inflation? Arch Intern Med. 2003;163:1637–41.
22. Fryback DG, Dasback EJ, Klein R, Klein BEK, Peterson K, Martin PA. The Beaver Dam health outcomes study: initial catalog of health-state quality factors. Med Decis Making. 1993;13:89–102.
23. Chang RW, Pellissier JM, Hazen GB. A cost-effectiveness analysis of total hip arthroplasty for osteoarthritis of the hip. JAMA. 1996;275(11):858–65.
24. Mark DB, Hlatky MA, Califf RM, et al. Cost effectiveness of thrombolytic therapy with tissue plasminogen activator as compared with streptokinase for acute myocardial infarction. N Engl J Med. 1995;332(21):1418–24.
25. Gafni A, Birch S. Inclusion of drugs in provincial drug benefit programs: Should "reasonable decisions" lead to uncontrolled growth in expenditures? Can Med Assoc J. 2003;168(7):849–51.
26. Zethraeus N. Willingness to pay for hormone replacement therapy. Health Econ. 1998;7(1):31–8.
27. Exchange rates lookup. Available at: http://www.x-rates.com/cgi-bin/hlookup.cgi. Accessed 6 September 2002.
28. Cunningham SJ, Hunt NP. Relationship between utility values and willingness to pay in patients undergoing orthognathic treatment. Community Dent Health. 2000;17(2):92–6.
29. Lundberg L, Johannesson M, Silverdahl M, Hermansson C, Lindberg M. Quality of life, health-state utilities and willingness to pay in patients with psoriasis and atopic eczema. Br J Dermatol. 1999;141(6):1067–75.
30. Blumenschein K, Johannesson M. Relationship between quality of life instruments, health state utilities, and willingness to pay in patients with asthma. Ann Allergy Asthma Immunol. 1998;80(2):189–94.
31. Lloyd AJ. The extent of patients' understanding of the risk of treatments. Qual Health Care. 2001;10(Suppl 1):i14–8.
32. Kaplan RM, Sieber WJ, Ganiats TG. The quality of well-being scale: comparison of the interviewer-administered version with a self-administered questionnaire. Psychol Health. 1997;12:783–91.
33. Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems: health utilities index. Pharmacoeconomics. 1995;7:490–502.
34. EuroQol—a new facility for the measurement of health-related quality of life. The EuroQol Group. Health Policy. 1990;16(3):199–208.
35. Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. J Chronic Dis. 1978;31:697–704.
36. Boyd NF, Sutherland HJ, Heasman KZ, Tritchler DL, Cummings BJ. Whose utilities for decision analysis? Med Decis Making. 1990;10(58):67.
37. Najman J, Levine S. Evaluating the impact of medical care and technologies on the quality of life: a review and critique. Soc Sci Med. 1981;15F:105–15.
38. Epstein AM, Hall JA, Tognetti J, Son LH, Conant L. Using proxies to evaluate quality of life. Med Care. 1989;27:S91–8.
39. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med. 1999;48(11):1507–15.